



# Introdução ao Aprendizado de Máquina

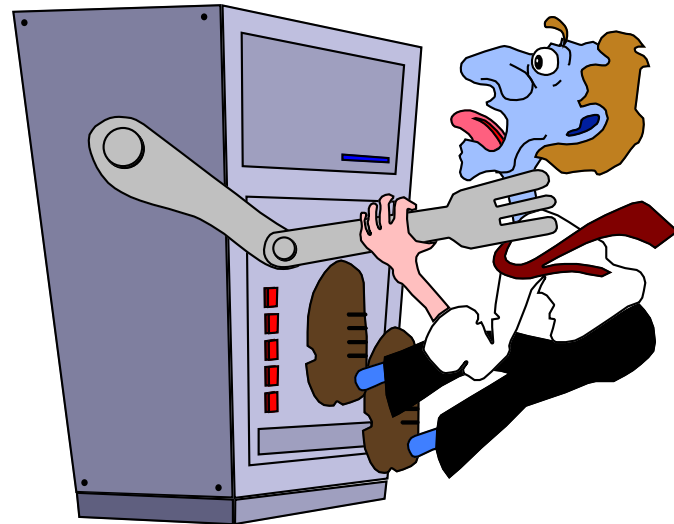
---

- Disciplina: Tópicos: Aprendizado de Máquina
- Profa. Dra. Adriane Beatriz de Souza Serapião
- Ciências da Computação – UNESP – Rio Claro

# O que é Inteligência Artificial?

***“Uma área de pesquisa que investiga formas de habilitar o computador a realizar tarefas nas quais, até o momento, o ser humano tem um melhor desempenho”.***

*Elaine Rich*





# Inteligência Artificial

---

“Tão logo algum problema de IA é resolvido ele não é mais considerado um problema da área de IA...”

Chuck Thorpe  
CMU, Robotics Institute, 2000



# Definições Adicionais

---

- Conjunto de técnicas para a construção de máquinas “inteligentes”, capazes de resolver problemas que requerem inteligência humana. (Nilsson)
- Ramo da Ciência da Computação dedicado à automação de comportamento inteligente. (Luger e Stubble)
- Tecnologia de processamento de informação que envolve raciocínio, aprendizado e percepção. (Winston)

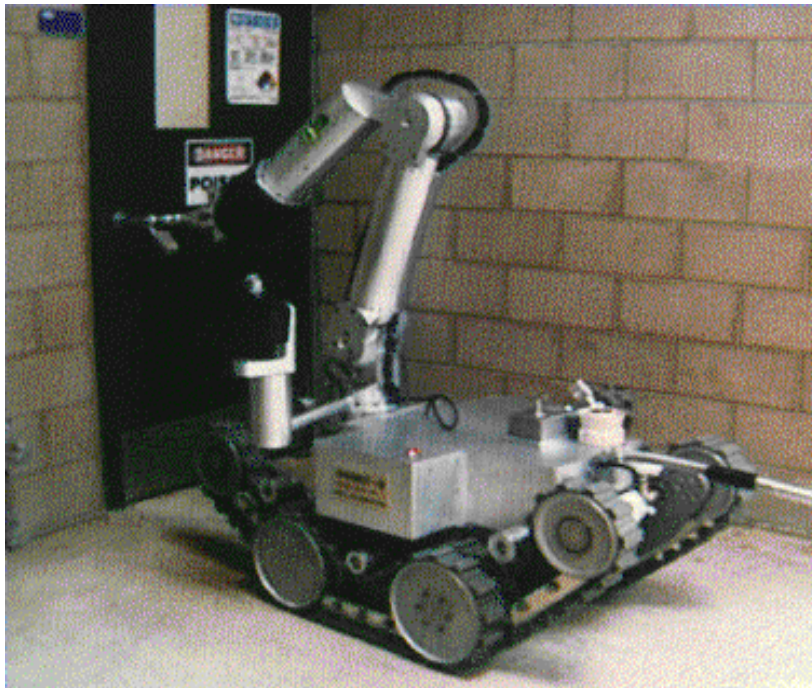
# Produção de jogos e histórias interativas

- Como modelar o ambiente físico e o comportamento/personalidade dos personagens?
- Como permitir uma boa interação com usuário?



# Controle de robôs

- Como obter navegação segura e eficiente, estabilidade, manipulação fina e versátil?
- E no caso de ambientes dinâmicos e imprevisíveis?



# Busca de informação na Web

- Como localizar a informação relevante?



[MIT Artificial Intelligence Lab](#) - [ [Traduzir esta página](#) ]

... The **Artificial Intelligence** Laboratory has been an active entity at MIT in one form or another since at least 1959. Our goal is ...

Descrição: Aiming to understand the nature of **intelligence**, to engineer systems that exhibit such **intelligence**...

Categoria:

[Computers](#) > [Artificial Intelligence](#) > [Academic Departments](#)

[www.ai.mit.edu/](http://www.ai.mit.edu/) - 22k - [Em cache](#) - [Páginas Semelhantes](#)

Links Patrocinados

[Will robots ever think?](#)

Here are the main models of cognition, so judge for yourself.

[www.smithsrisca.demon.co.uk/](http://www.smithsrisca.demon.co.uk/)

Interesse:

[Veja a sua mensagem](#)

# Previsão

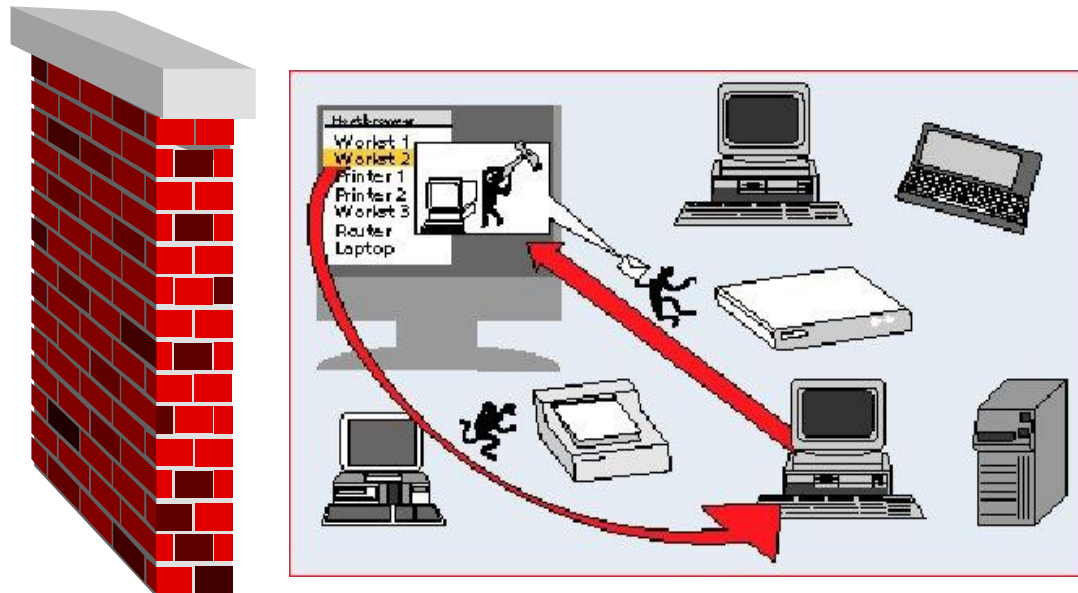
- Como prever o valor do dólar (ou o clima) amanhã?
- Que dados são relevantes? Há comportamentos recorrentes?





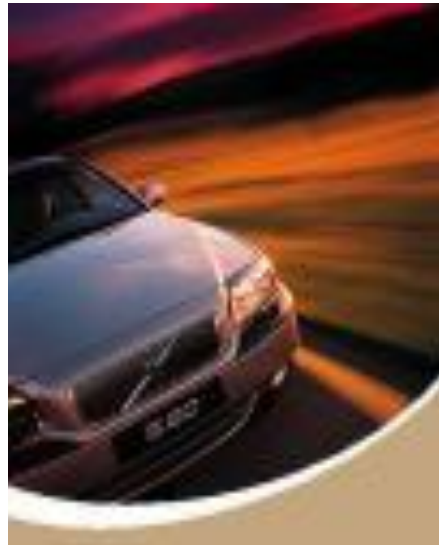
# Detecção de Intrusão e Filtragem de Spam

- Como saber se uma mensagem é lixo ou de fato interessa?
- Como saber se um dado comportamento de usuário é suspeito e com lidar com isto?



# Sistemas de Controle

- Como breicar o carro sem as rodas deslizarem em função da velocidade, atrito, etc.?
- Como focar a câmera em função de luminosidade, distância, etc.?
- Como ajustar a temperatura em da quantidade de roupa, fluxo de água, etc.?



# Interface

- Como dar ao usuário a ajuda de que ele precisa?
- Como interagir (e quem sabe navegar na web) com celular sem ter de digitar (hands-free)?





# O que estes problemas têm em comum?

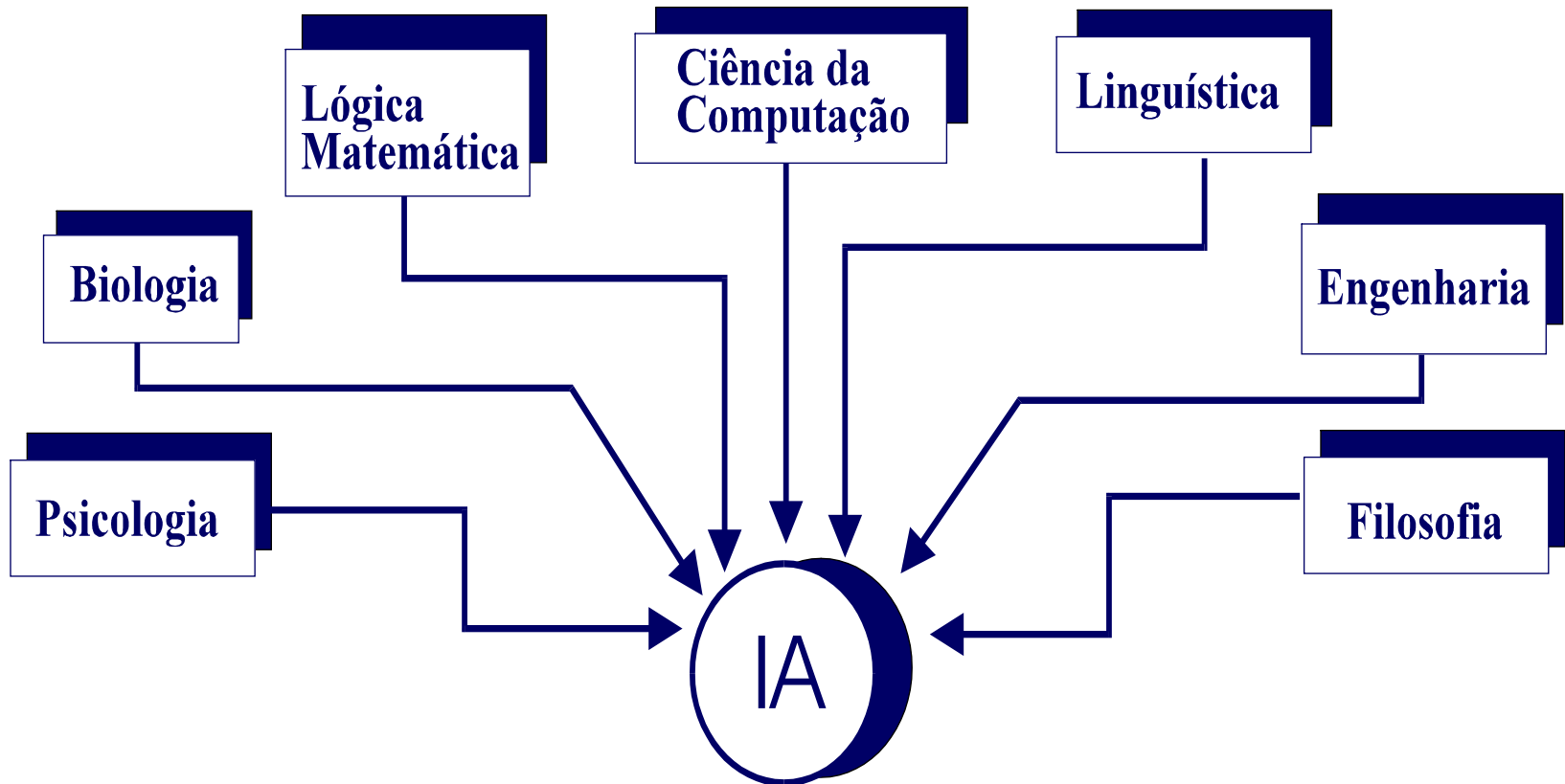
---

- Grande complexidade (número, variedade e natureza das tarefas).
- Não há “solução algorítmica”, mas existe conhecimento.
- Modelagem do comportamento de um ser inteligente (conhecimento, aprendizagem, iniciativa, etc.).



# Áreas de Apoio para IA

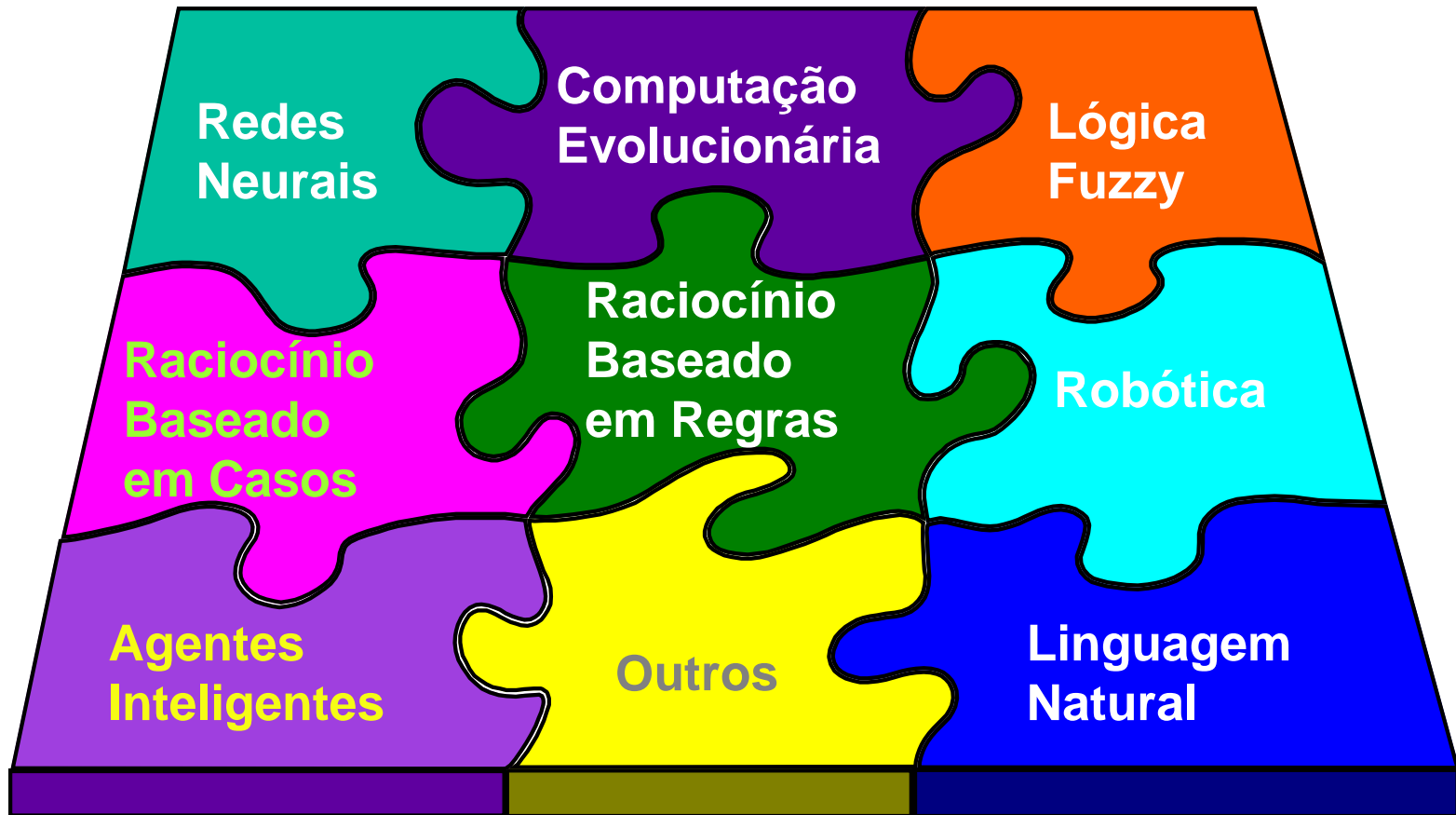
---





# Sub-Áreas da IA

---





# Contexto Histórico

---

“IA é considerada polêmica porque desafia a ideia da unicidade do pensamento humano, da mesma forma que Darwin desafiou a unicidade da origem dos seres humanos.”

Herbert A. Simon  
CMU, 2000

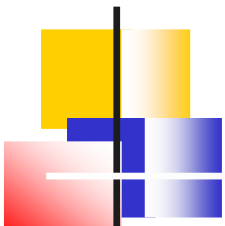


# Objetivo da disciplina

---

- Estudo de métodos e técnicas envolvendo o processo de criação de sistemas computacionais que apresentem características de aprendizagem.
  - Fornecer uma visão geral da área de aprendizagem de máquina.
  - Estudar métodos e técnicas de aprendizagem de máquina.





# Motivação

---

- Desde que os computadores foram inventados temos nos perguntado:  
“Eles são capazes de aprender?”
- Infelizmente ainda não sabemos como fazer computadores aprenderem de uma maneira similar a como os humanos aprendem.
- Entretanto, foram desenvolvidos algoritmos que são eficientes em certos tipos de tarefas de aprendizagem e um entendimento teórico de aprendizagem está começando a surgir.



# Motivação

---

- Em geral, é difícil articular o conhecimento que precisamos para construir um sistema de IA.
- Na verdade, algumas vezes, não temos nem este conhecimento.
- Em alguns casos, podemos construir sistemas em que eles mesmos aprendem o conhecimento necessário.



# O que é Aprendizado?

---

- Aprender fatos por meio de observação e exploração.
- Melhorar habilidades motoras/cognitivas por meio de prática.
- Organizar novo conhecimento em representações efetivas e gerais.



# Inteligência vs. aprendizado

---

- Aprendizado é a chave da superioridade da inteligência humana.
- Para que uma máquina tenha *comportamento inteligente*, deve-se aumentar a *capacidade de aprendizado*.



# Inteligência vs. aprendizado

---

- O ser humano está **pré-programado para o aprendizado**; aprende ampliando o alcance do conhecimento que já possui, através de reordenações sucessivas.
- O computador não possui o programa inicial para procurar por informações e realizar aprendizado em geral.
- **Paradigmas e técnicas de AM** possuem um alvo bem mais limitado do que o aprendizado humano.



# O que é Aprendizagem de Máquina?

---

- Área de estudo que fornece aos computadores a habilidade de aprender sem serem explicitamente programados [Arthur Samuel (1959)].
- **Definição bem formada:** *A computer program is said to learn from experience  $A$  with respect to some task  $T$  and some performance measure  $P$ , if its performance on  $T$ , as measured by  $P$ , improves with experience*



# Aprendizado de Máquina

---

- **Principal preocupação:**

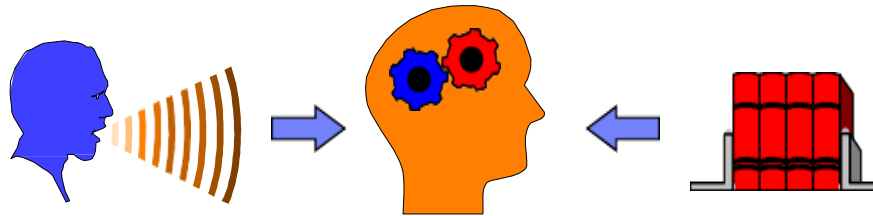
- Construção de programas de computador que melhoram seu desempenho por meio de experiência.

- **Técnicas orientadas a dados:**

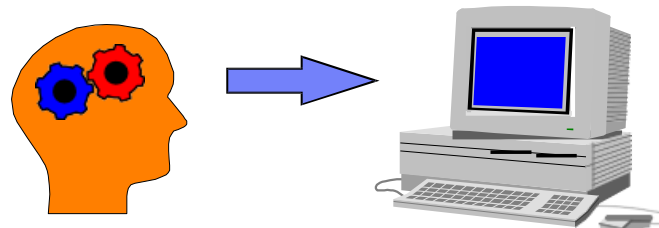
- Aprendem automaticamente a partir de grandes volumes de dados.
- Geração de hipóteses a partir dos dados.

# Objetivos do aprendizado de máquina


- Um melhor entendimento dos mecanismos de aprendizado humano.



- Automação da aquisição do conhecimento.







# Exemplos aprendizagem de máquina

---

- A partir de informações sobre pacientes relativas a gravidez aprender a prever classes de futuros pacientes de alto risco que devem fazer cesárea.
- Análise de risco de crédito: prever clientes mal pagadores.
- Prever comportamento de compra de clientes.
- Recomendar filmes para clientes.
- Etc.



# Multidisciplinaridade da Aprendizagem de Máquina

---

- Inteligência Artificial
- Estatística
- Teoria da Informação
- Teoria de Controle
- Filosofia
- Psicologia
- Neurobiologia
- ...



# Paradigmas de AM

---

- Simbólico
- Baseado em exemplos
- Estatístico
- Conexionista
- Evolutivo



# Paradigmas de AM – Simbólico

---

- Explora representações de estruturas gráficas ou lógicas, no lugar de métodos estatísticos ou numéricos.
- Descrições simbólicas representam um conhecimento de alto nível.
- As representações simbólicas estão tipicamente na forma de alguma expressão lógica, árvore de decisão, regras de produção ou rede semântica.



# Paradigmas de AM – Baseado em exemplos

---

- Exemplos de treinamento são armazenados.
- Casos nunca vistos são classificados através de casos similares conhecidos.
- Classificação de um caso é lembrar de um caso similar cuja classe é conhecida e assumir que o novo caso terá a mesma classe.



# Paradigmas de AM – Estatístico

---

- Decisões tomadas através de raciocínio sobre probabilidades dos dados.
- Como regra geral, técnicas estatísticas tendem a focar tarefas em que todos os atributos têm valores contínuos ou ordinais.
- Muitas técnicas são paramétricas, assumindo alguma forma de modelo, e então encontrando valores apropriados para os parâmetros do modelo a partir de dados.



# Paradigmas de AM – Conexionista

---

- Estudo de Redes Neurais Artificiais foi inspirado em parte na observação de que sistemas de aprendizado biológico são compostos de redes muito complexas de neurônios interconectados.
- Redes Neurais Artificiais são redes construídas a partir de conjuntos de unidades simples altamente interconectadas, daí o nome *conexionismo*.



# Paradigmas de AM – Evolutivo

---

- Este paradigma possui uma analogia direta com a teoria de Darwin, onde sobrevivem os mais bem adaptados ao ambiente.
- Um classificador evolutivo consiste em uma população de elementos de classificação que competem para fazer a predição; elementos que possuem uma performance fraca são descartados, enquanto os elementos mais fortes proliferam, produzindo variações de si mesmos.

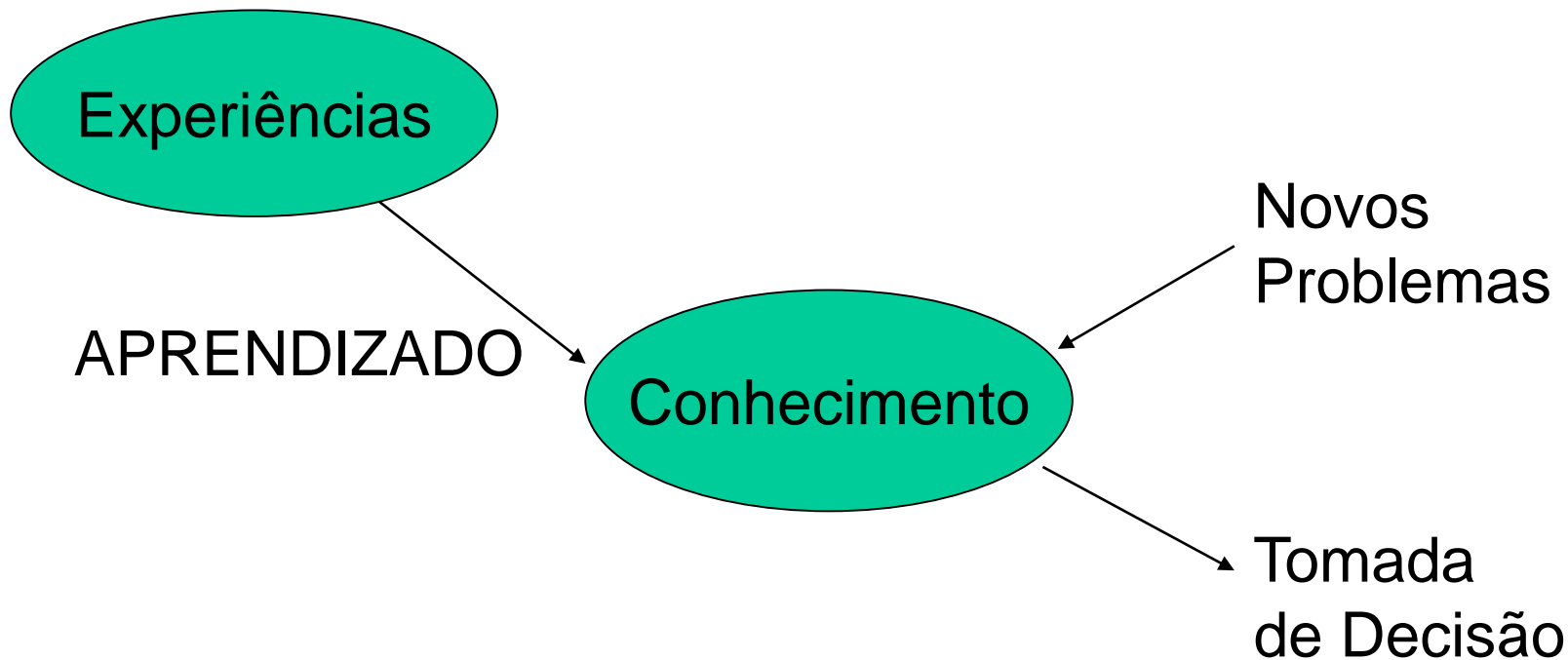




# Aprendizado de Máquina

---

- Conhecimento adquirido é usado para auxiliar a *tomada de decisões*.





# Argumentos dedutivos vs. indutivos

---

## ■ Argumento dedutivo

- *Nenhum dos alunos gosta de Inteligência Artificial*  
*Francisco é um aluno*
- $\Rightarrow$  *Francisco não gosta de Inteligência Artificial*

## ■ Argumento indutivo

- *Nenhum dos alunos que foram entrevistados gosta de Inteligência Artificial*
- $\Rightarrow$  *Nenhum aluno gosta de Inteligência Artificial*

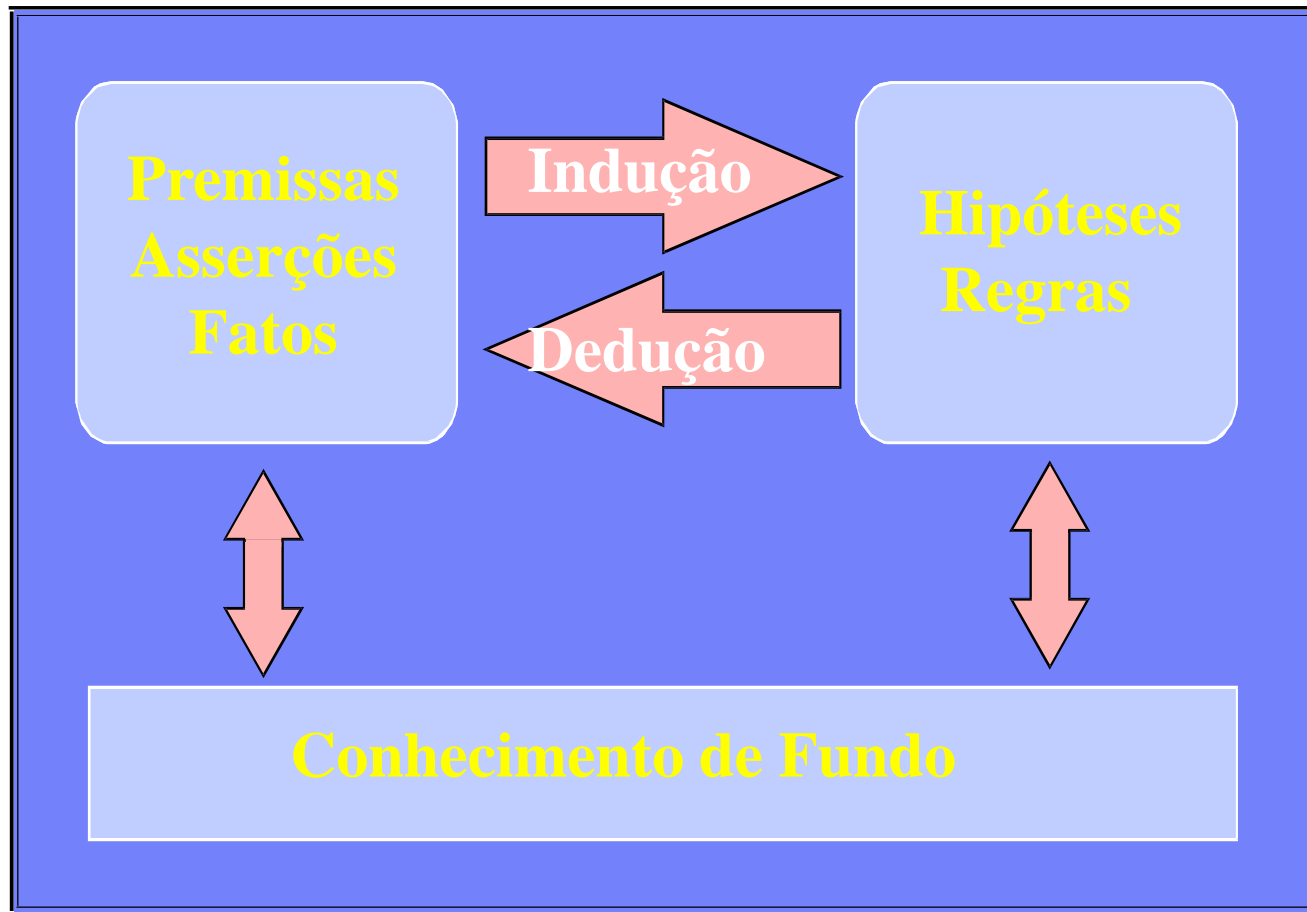


# Argumentos dedutivos vs. indutivos

---

Argumentos dedutivos	Argumentos indutivos
Se todas as premissas são verdadeiras, conclusão é verdadeira	Se todas as premissas são verdadeiras, a conclusão é provavelmente verdadeira, mas não necessariamente verdadeira (à exceção dos argumentos matemáticos indutivos).
Toda a informação do conteúdo fatural da conclusão já está, pelo menos, implicitamente nas premissas.	A conclusão contém informação que não está nem implicitamente nas premissas.

# Relação entre dedução e indução





# Aprendizado de Máquina

---

- Conhecimento é adquirido através de *inferência indutiva* (*indução*).
- Para discussão:
  - Qual a diferença entre *dedução* e *indução*?
    - Dedução: gera fatos a partir de premissas.
    - Indução: gera premissas a partir de fatos.



# Aprendizado por indução

---

- Dependendo dos fatos necessários para realizar o aprendizado fornecidos por uma fonte externa ou observados pelo sistema aprendiz, pode-se distinguir dois tipos diferentes de estratégias de aprendizado
  - Aprendizado por **observação e descoberta.**
  - Aprendizado por **exemplos.**

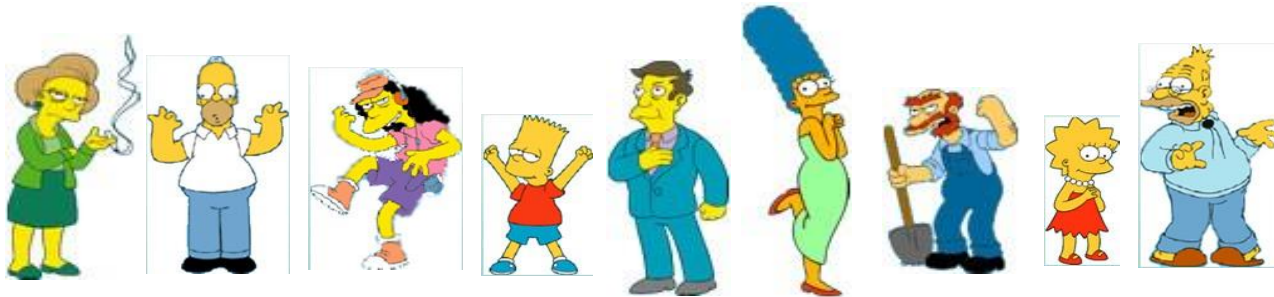
# Aprendizado por indução

- Aprendizado por observação e descoberta
  - Não existe o professor.
  - O aprendiz analisa entidades fornecidas ou observadas e tenta determinar se alguns subconjuntos dessas entidades podem ser agrupados em certas classes de maneira útil.
  - Aprendizado Não-Supervisionado.



# Agrupamento

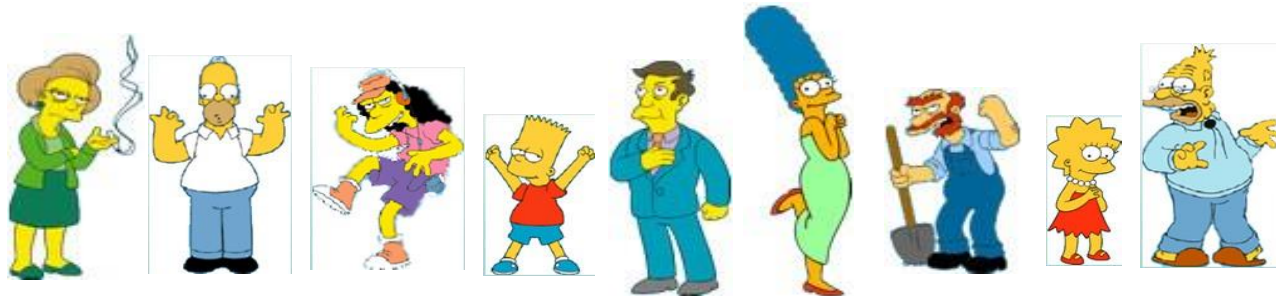
- Como agrupar estes objetos?



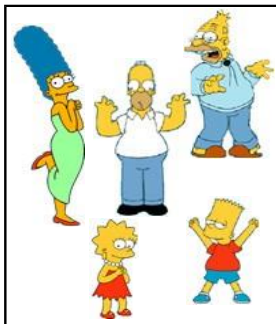


# Agrupamento

- Como agrupar estes objetos?



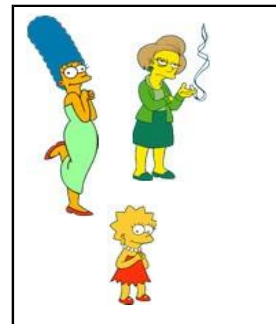
- O agrupamento é subjetivo



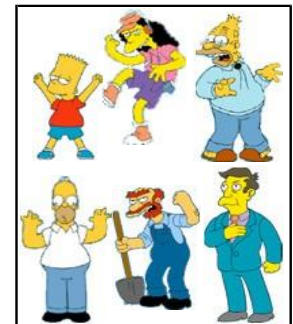
Os Simpsons



Empregados da escola



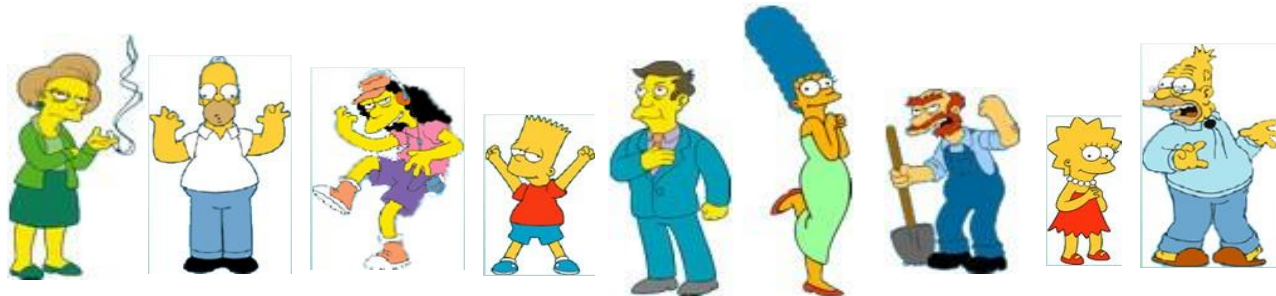
Mulheres



Homens

# Agrupamento

- Como agrupar estes objetos?



- Qual seu critério de agrupamento?
  - Similaridade em algum nível.

# O que é similaridade?

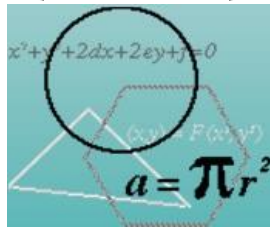
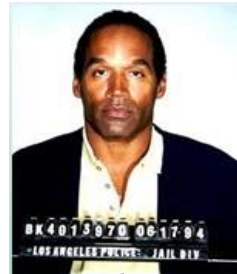
- A qualidade, caráter ou condição das coisas similares.  
(*Dicionário Houaiss*)



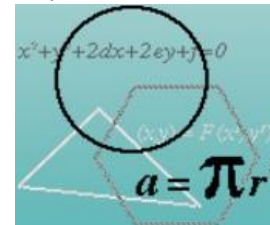
Similaridade é  
difícil de definir,  
mas...  
Reconhecemos  
quando a vemos!

# Como medir similaridade?

- Função da distância ou similaridade.



características =  
cor da pele, altura  
distância = 0.2



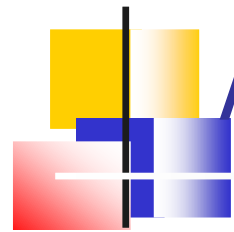
características =  
sobrenome, idade  
distância = 0.4



# Aprendizado por indução

---

- Aprendizado por exemplos
  - Existe um professor que já tem conhecimento do conceito.
  - O professor ajuda na escolha dos exemplos.
  - O aprendiz induz a descrição de um conceito, formulando uma regra geral a partir de exemplos e contra-exemplos.
  - TAREFA: determinar a descrição geral de um conceito.
  - Aprendizado Supervisionado.



# Aprendizado por indução

---

- Aprendizado por exemplos
  - Nesta estratégia de aprendizado, o sistema aprendiz induz a descrição do conceito ou classe através de processos de generalização e especialização realizados sobre exemplos e, opcionalmente, contra-exemplos do conceito.



# Aprendizado por indução

---

- **Inferência indutiva** é um dos principais meios para a aquisição de novos conhecimentos e previsão de eventos futuros.
- **Observações permitem descobrir regras e procedimentos.**
- Deve-se ter cuidado com o número de observações e a relevância dos dados.



# Características do aprendizado indutivo

---

- Dois tipos de aprendizado:
  - 1. Aprendizado incremental
    - Em geral, no aprendizado incremental, o processo de aprendizado procede através de uma sequência de hipóteses,  $H_1, H_2, \dots$  etc., sobre o conceito que está sendo aprendido. Quando um exemplo é processado, a hipótese corrente é atualizada, se necessário, resultando na próxima hipótese.





# Características do aprendizado indutivo

---

- Dois tipos de aprendizado:
  - 2. Aprendizado não-incremental
    - Necessita de que todos os exemplos de treinamento, simultaneamente, estejam disponíveis para que seja induzido um conceito.
    - É vantajoso usar esses algoritmos para problemas de aprendizado onde todos os exemplos estão disponíveis e, provavelmente, não irão ocorrer mudanças.



# Características do aprendizado indutivo - LD

---

- Linguagens de descrição:
  - ◆ LD de instâncias  $L_\varepsilon$  (Exemplos/Objetos)
  - ◆ LD de conceitos  $L_H$  (Hipóteses)
  - ◆ LD da teoria do domínio  $L_K$  (conhecimento de fundo)



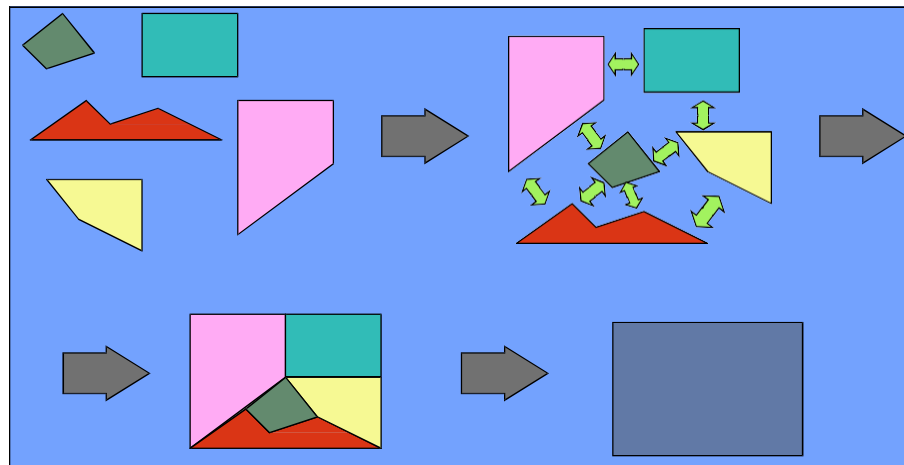
# Características do aprendizado indutivo - LD

---

- LD de instâncias  $L_\varepsilon$  (Exemplos/Objetos)
  - Qualquer que seja o tipo de aprendizado, é **necessário uma linguagem para descrever objetos** (ou possíveis eventos) e uma linguagem para descrever conceitos.
  - Em geral, é possível distinguir dois tipos de descrições para objetos: **estrutural** e **atributos**.

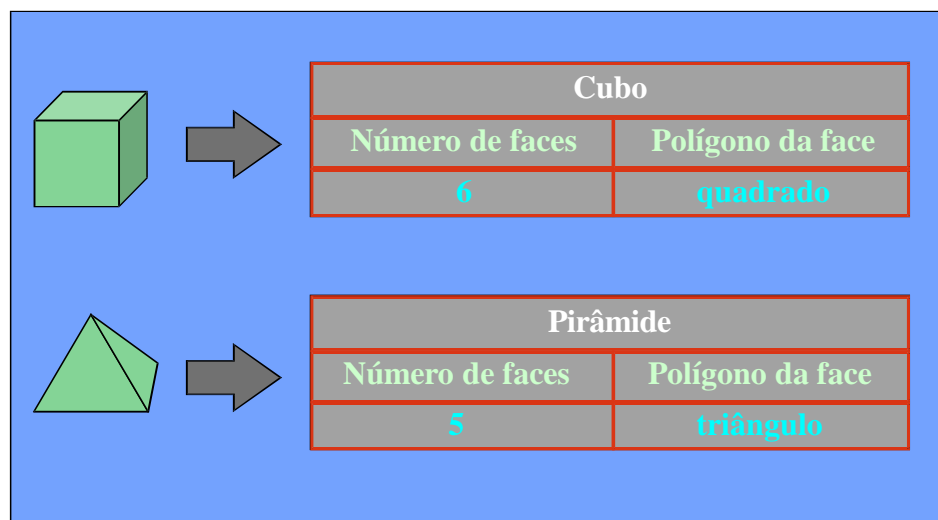
# Características do aprendizado indutivo - LD

- LD de instâncias  $L_\varepsilon$  (Exemplos/Objetos)
- 1. Descrições estruturais: um objeto é descrito em termos de seus componentes e a relação entre eles.



# Características do aprendizado indutivo - LD

- LD de instâncias  $L_{\varepsilon}$  (Exemplos/Objetos)
- 2. Descrições de atributos: um objeto é descrito em termos de suas características globais como um vetor de valores de atributos.





# Características do aprendizado indutivo - LD

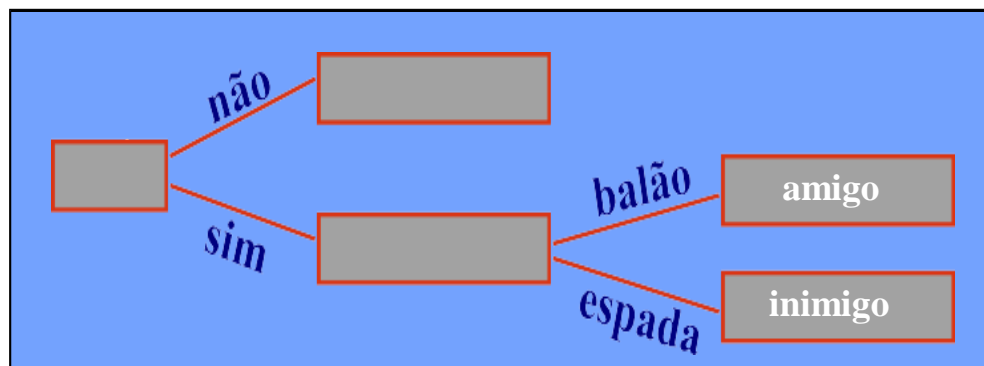
---

- LD de conceitos  $L_H$  (Hipóteses)
- Formalismos frequentemente usados em AM para descrever conceitos são:
  - regras *se-então* (*if-then*) para representar conceitos

*Se Nublado ou Chovendo  
então Levar\_Guarda-Chuva*

# Características do aprendizado indutivo - LD

- LD de conceitos  $L_H$  (Hipóteses)
- Formalismos frequentemente usados em AM para descrever conceitos são:
  - árvores de decisão para representar conceitos





# Características do aprendizado indutivo - LD

---

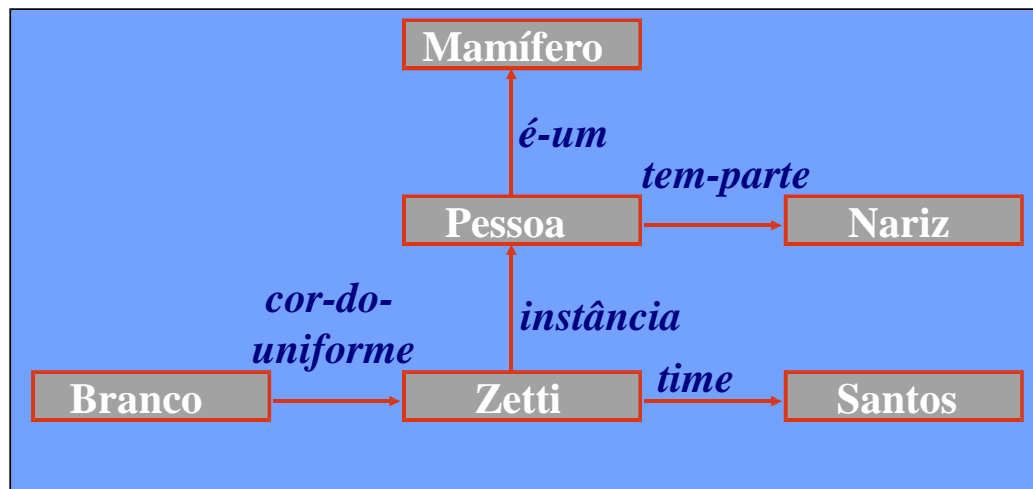
- LD de conceitos  $L_H$  (Hipóteses)
- Formalismos frequentemente usados em AM para descrever conceitos são:
  - lógica de predicados

```
filha(X, Y) <-- mulher(X), pais(X, Y).
```



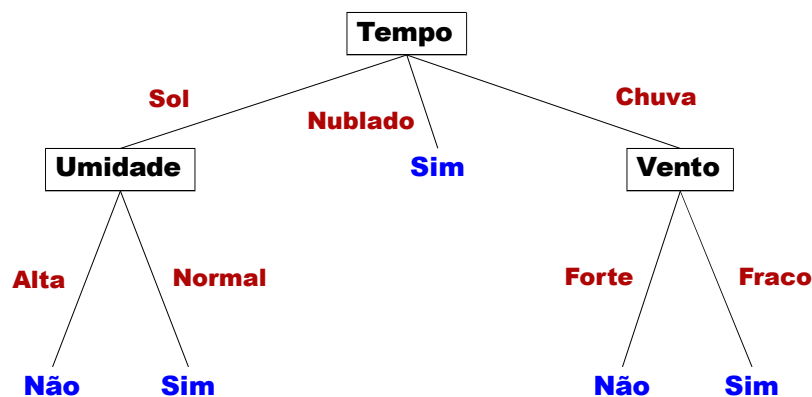
# Características do aprendizado indutivo - LD

- LD de conceitos  $L_H$  (Hipóteses)
- Formalismos frequentemente usados em AM para descrever conceitos são:
  - redes semânticas



# Características do aprendizado indutivo - LD

- LD de conceitos  $L_H$  (Hipóteses)
- Qual o formalismo?





# Características do aprendizado indutivo - LD

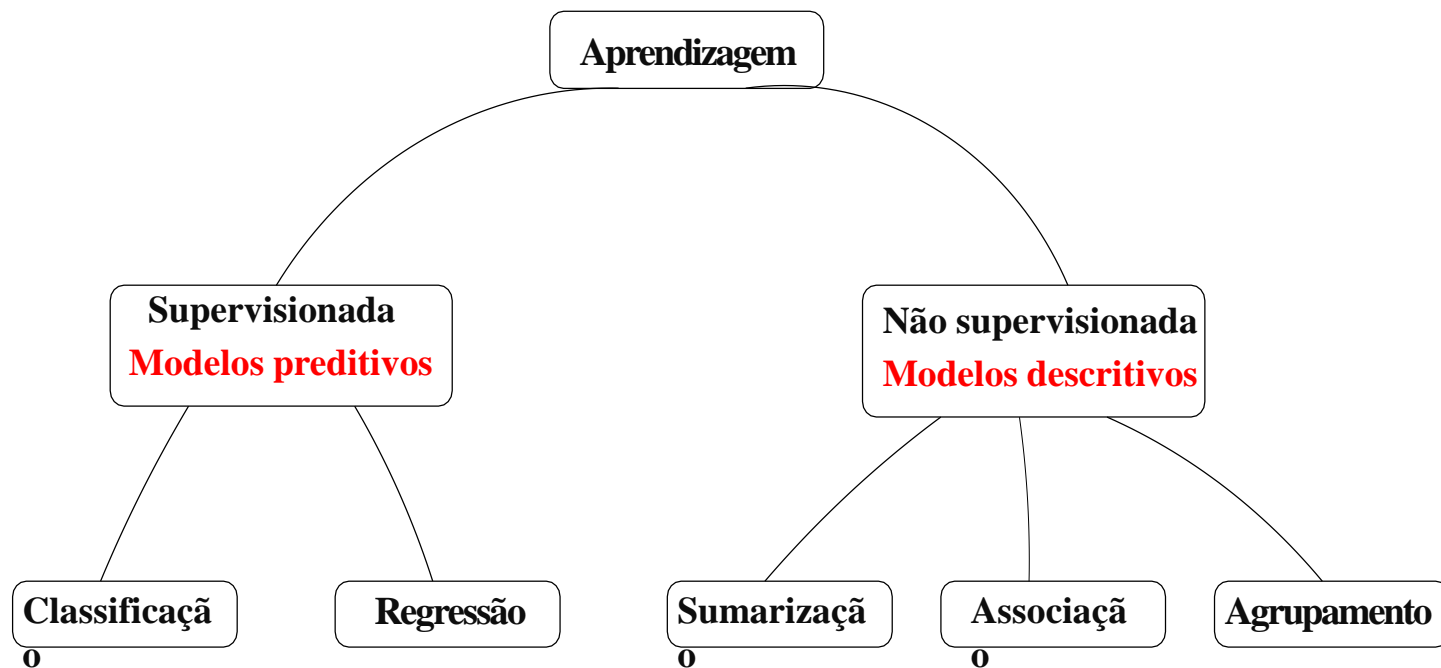
---

- LD da teoria do domínio  $L_K$  (conhecimento de fundo)
  - Conhecimento de fundo: constituído por algum conhecimento relevante do domínio do problema.
  - Exemplo do viajante na Itália:
    - A generalização de que todos os italianos falam italiano é sustentada pela regularidade mais geral de que em um dado país a maioria da população fala a mesma língua; por outro lado, não é assumido que todos os italianos são chamados de Giuseppe devido à regularidade mais geral de que a maioria dos grupos sociais utilizam nomes diversos para diferentes indivíduos.



# Aprendizado de Máquina

---






# Aprendizado de Máquina

---


- Modos de aprendizado
  - Aprendizado Supervisionado:
    - Auxílio de um professor que nos diz algo a respeito dos objetos que observamos.
      - Na prática, relaciona-se com resolução de problemas de classificação e regressão.
  - Aprendizado Não-Supervisionado:
    - Mesmo sem um professor somos capazes de identificar padrões nos objetos que observamos.
      - Na prática, relaciona-se com problemas de agrupamento e geração de regras de associação.



# Aprendizado de Máquina Supervisionado

---

- Classificação
  - Associar objetos a uma categoria ou classe.
    - E.g., diagnóstico de pacientes, classificação risco de um cliente, classificação de documentos,...
  - Classificação é feita com base nos atributos dos objetos.
    - E.g., diagnóstico de um paciente é feito com base nos sintomas observados e exames realizados.
  - Aprendemos a classificar melhor com o tempo à medida que observamos novos exemplos.



# Aprendizado de Máquina Supervisionado

---

- Regressão
  - Associar objetos a valores numéricos.
    - E.g., previsão de índices da bolsa de valores, predição de custo de desenvolvimento de software,...
  - Similar à classificação, porém atributo alvo é numérico.




# Exemplos de aprendizagem supervisionada

---

- Estimar o preço de uma casa.
  - atributos: tamanho, posição geográfica, material.
  - classe: preço (regressão).
- Determinar se uma pessoa tem câncer benigno ou maligno.
  - atributos: tamanho do tumor, formato do tumor, idade do paciente.
  - classe: tumor benigno ou tumor maligno (classificação).





# Exemplos de aprendizagem supervisionada

---

- Determinar se é um texto publicado em uma rede social é inadequado ou não.
  - Atributos: quantidade de palavras encontradas no texto, quantidade de palavras proibidas encontradas no texto, quantidade de textos já criados pelo usuário, idade do usuário no sistema, quantidade de textos criados pelo usuário e moderados, ...
  - classe: texto adequado ou não (classificação).
  - **classes**: texto adequado, texto inadequado, texto com propaganda (classificação com múltiplas



# Exemplo de conjunto de dados com classe

---

Idade	Miopia	Astigmat.	Lacrimaj.	Lentes
jovem	míope	não	reduzido	<b>nenhuma</b>
jovem	míope	não	normal	<b>fraca</b>
jovem	míope	sim	reduzido	<b>nenhuma</b>
jovem	míope	sim	normal	<b>forte</b>
...	...	...	...	...
adulto	míope	não	reduzido	<b>nenhuma</b>



# Exercícios

---

- Que problema deve ser tratado como problema de regressão e que problema deve ser tratado como problema de classificação?
- A sua empresa possui 1.000 itens idênticos em estoque. Você quer prever quantos destes itens serão vendidos nos próximos três meses.
- Você quer examinar clientes seus e para cada um decidir se ele irá pagar todo o financiamento ou não.



# Aprendizado de Máquina Não-Supervisionado

---

- Agrupamento
  - Identificar grupos de objetos similares entre si e diferentes de objetos de outros grupos.
    - E.g., Identificar grupos de genes similares, agrupar resultados de engenhos de busca,...
  - Nos seres humanos, esse tipo de tarefa é realizada mesmo antes do desenvolvimento da linguagem.



# Aprendizado de Máquina Não-Supervisionado

---

- Regras de Associação
  - Identificar relacionados frequentes entre variáveis que descrevem objetos.
    - E.g., análise “market basket”,...



# Exemplos de aprendizagem não supervisionada

---

- Dado conjuntos de itens adquiridos na mesma compra, identificar padrões de compra.
- Identificar padrões de navegação em sites.
- Agrupar notícias semelhantes publicadas por várias fontes de informação.
- Numa rede social, identificar sub-grupo de pessoas.



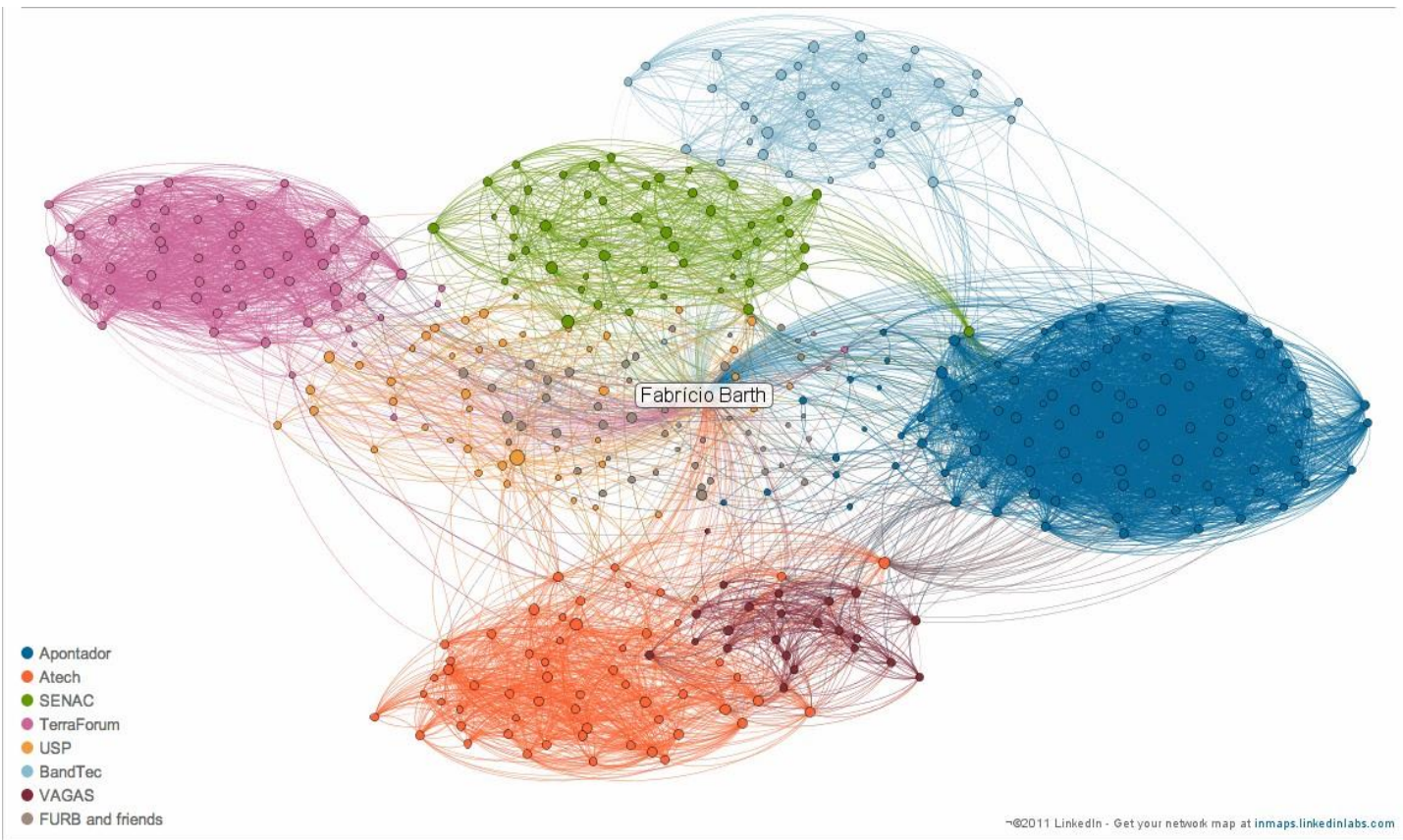
# Exemplos de aprendizagem não supervisionada

---

Tabela 1: Exemplo de tabela com as transações dos usuários

usuário	<i>categoria</i> <sub>1</sub>	<i>categoria</i> <sub>2</sub>	<i>categoria</i> <sub>3</sub>	...	<i>categoria</i> <sub><i>m</i></sub>
<i>user</i> <sub>1</sub>	0	2	0	...	1
<i>user</i> <sub>2</sub>	1	1	0	...	0
<i>user</i> <sub>3</sub>	2	0	1	...	0
<i>user</i> <sub>4</sub>	0	1	0	...	0
...	...	...	...	...	...
<i>user</i> <sub><i>n</i></sub>	1	1	0	...	1

# Exemplo de identificação de grupos em redes sociais







# Exercícios

---

- Quais dos problemas abaixo você iria resolver com uma abordagem não supervisionada de aprendizagem?
  - Dado e-mails rotulados como spam e não spam, desenvolver um filtro de spam.
  - Dado um conjunto de notícias encontradas na Internet, agrupá-las em conjunto de notícias que tratam do mesmo assunto.
  - Dado uma base de clientes, descobrir segmentos de clientes.
  - Dado uma base de pacientes diagnosticados com



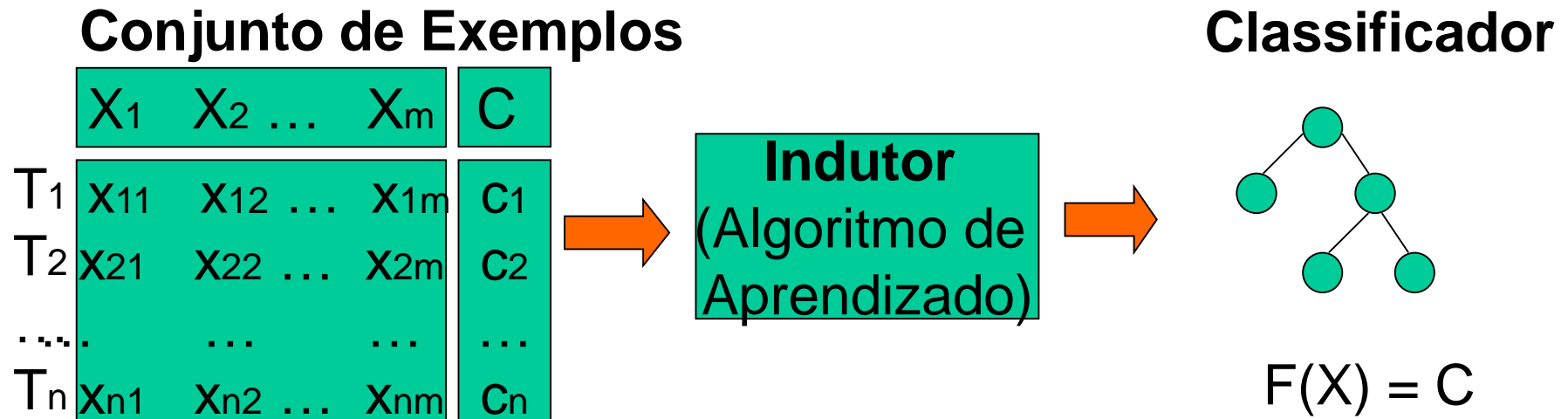
# Aprendizado de Máquina

## Conceitos

# Aprendizado de Máquina Supervisionado

## ■ Indutor

- Algoritmo que adquire conhecimento a partir de um conjunto de exemplos.





# Aprendizado de Máquina

## Supervisionado

---

- Exemplo (ou instância):
  - Tupla com atributos que descrevem um objeto de interesse + classe do exemplo.
    - E.g., dados de um paciente + doença
- Atributos descritores:
  - Característica de um exemplo usada para classificação.
- Atributo classe:
  - Atributo alvo da predição.



# Aprendizado de Máquina

## Supervisionado

---

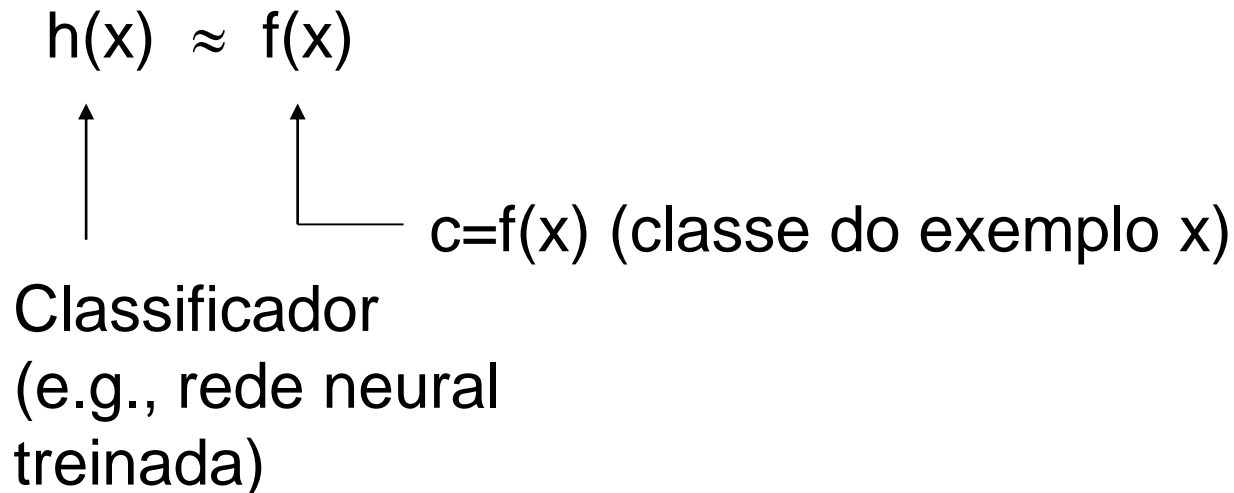
- Tipos de Atributos:
  - Numérico X categórico
    - E.g., Peso (Kg) X Classe social (A, B, C, ...)
  - Discreto X contínuo
    - E.g., Idade X Temperatura
  - Ordinal X nominal
    - E.g., Estatura (Alta, Baixa) X Cor (Azul, Verde)



# Aprendizado de Máquina Supervisionado

---

- Classificador (ou Hipótese ou Modelo)
- Resultado retornado pelo indutor (aproxima a *função real* de classificação)





# Aprendizado de Máquina Supervisionado


---

- Erro de Predição

- Taxa de erro de um classificador  $h$

$$erro(h) = \frac{1}{n} \sum_{i=1}^n \| c_i \neq h(x_i) \|$$

- Pode ser calculado durante treinamento e também em uma amostra de teste.



# Aprendizado de Máquina Supervisionado


---

- Erro Majoritário
  - Erro obtido com o classificador *default*
    - Instâncias a serem classificadas são sempre associadas à classe de maior frequência no treinamento.

$$erro\_maj(T) = 1 - \max_{j=1,\dots,k} distr(C_j)$$

- Limite abaixo do qual o erro de um classificador deve ficar.





# Aprendizado de Máquina Supervisionado

---

- Desbalanceamento das Classes
  - Ocorre quando uma classe ocorre na maioria dos exemplos.
    - Obviamente erro majoritário é baixo.



# Aprendizado de Máquina Supervisionado

---

- Ruído

- Imperfeições nos dados (tanto nos atributos descritores como nas classes):
  - Erros de coleta e preenchimento dos dados.
  - Falhas ou baixa qualidade nos instrumentos que registram os dados.
  - Aleatoriedade intrínseca dos dados.



# Aprendizado de Máquina Supervisionado


---

## ■ *Overfitting*

- Ajuste excessivo dos dados:
  - Generalização excessiva.
  - “Aprende” o ruído dos dados.
  - Baixo erro no treinamento, mas alto erro durante uso do classificador.

## ■ *Underfitting*

- Generalização insuficiente dos dados:
  - Alto erro tanto no treinamento e também no uso dos classificadores.



# Aprendizado de Máquina Supervisionado

---

- Qualidade dos Atributos
  - Irrelevantes:
    - Não têm relação com o atributo-alvo.
      - E.g., CPF e doença.
  - Redundantes:
    - São desnecessários quando colocados no contexto de outro atributo.
      - E.g., Classe social e renda mensal.




# Aprendizado de Máquina Supervisionado

---

- *Missing Values*

- Valores faltosos em um atributo.
- Pode ser ocasionado por erro:
  - E.g., Quebra de um equipamento em um dado intervalo de tempo.
- Mas algumas vezes contêm informação relevante:
  - E.g., Exame que um médico deixou de pedir.



# Aprendizado de Máquina Supervisionado

---

## ■ *Outliers*

- São dados específicos que diferem muito dos outros dados.
- Podem ser ocasionados por falhas de medição.
- Podem ser ocasionados por situações atípicas:
  - E.g., aumentos abruptos da bolsa de valores em momentos de crise.
  - E.g., fraude em cartão de crédito.



# Aprendizado de Máquina Supervisionado

---

- Existe uma grande diversidade de algoritmos de aprendizado.
- Tipos de algoritmos:
  - Árvores de Decisão e Regras.
  - Redes Neurais Artificiais.
  - Máquinas de Vetores Suporte.
  - Aprendizado Baseado em Instâncias.
  - Aprendizado Bayesiano.



# Sistema de aprendizado de máquina

---

<b><i>Modo de aprendizado</i></b>	<b><i>Paradigmas de Aprendizado</i></b>	<b><i>Linguagens de descrição</i></b>	<b><i>Formas de aprendizado</i></b>
-Supervisionado -Não Supervisionado	-Simbólico  -Estatístico  -Baseado em exemplos  -Conexionista  -Evolutivo	-Instâncias ou exemplos  -Conceitos aprendidos ou hipóteses  -Teoria de Domínio ou Conhecimento de Fundo	-Incremental  -Não Incremental



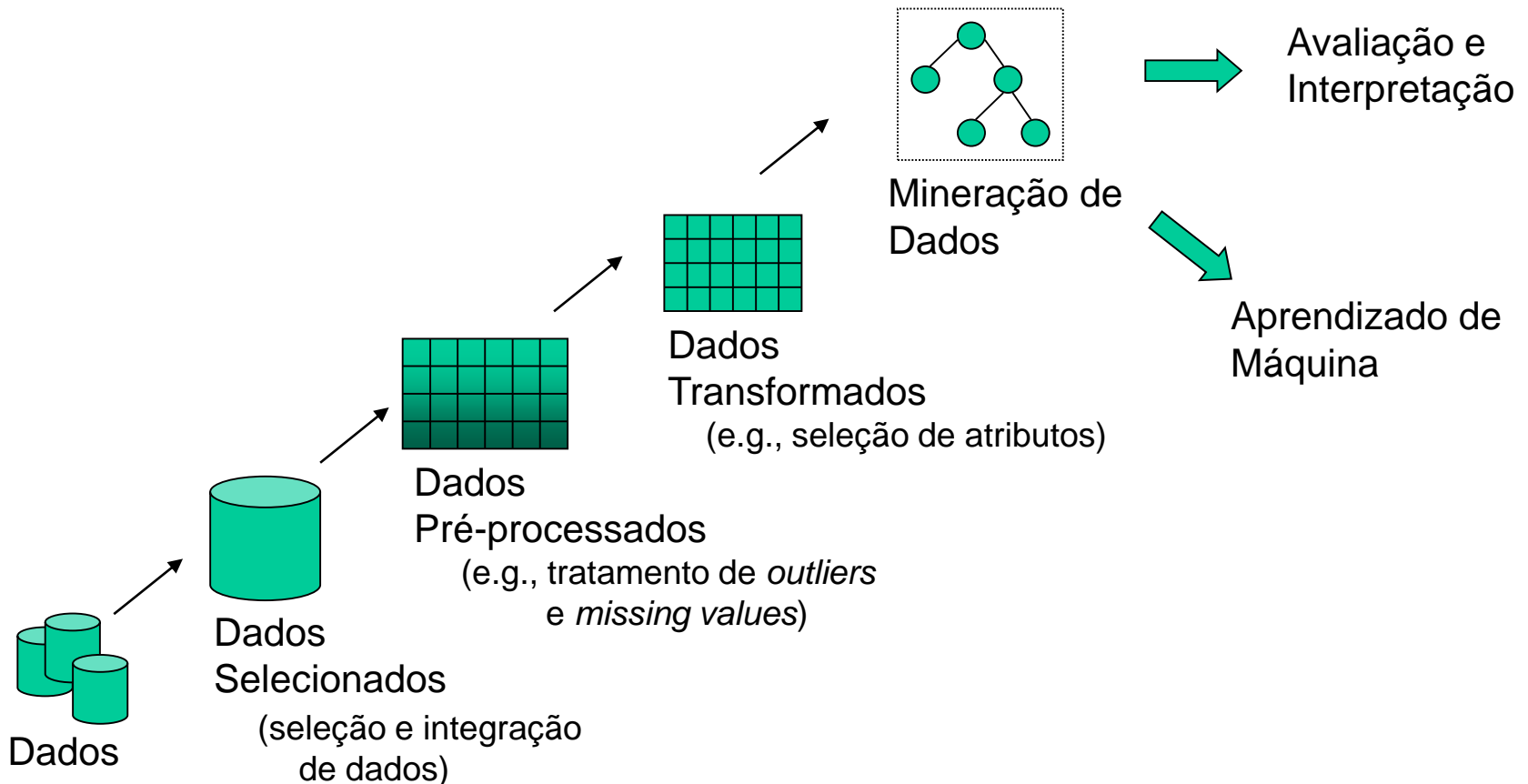


Aprendizado de Máquina

Aplicações

# Aprendizado de Máquina e KDD

## ■ KDD (*Knowledge Discovery in Databases*)






# Aprendizado de Máquina

---


- Biometria e reconhecimento de imagens.
- Aplicações em Engenharia:
  - Diagnóstico de falhas de transformadores, previsão de vazão hidrográfica, monitoramento de falhas em reatores,...
- Finanças e Marketing:
  - *Market basket analysis*, análise de fidelidade de clientes, análise de crédito, mineração de dados corporativos,...



# Aprendizado de Máquina e Mineração de Texto

---


- Classificação de documentos de texto:
  - Eg., Anti-Spam.
- Agrupamento de documentos de texto:
  - Visualização de bases de documentos.
- Extração de informação.



# Aprendizado de Máquina e Engenharia de Software

---

- Predição de qualidade de software.
- Predição de custo de software:
  - Desenvolvimento e teste.
- Predição de falhas.



# Aprendizado de Máquina e Bioinformática

---

- Agrupamento de dados de expressão gênica.
- Identificação de regiões promotoras.
- Identificação de início de sítios de tradução.



# Aprendizado de Máquina - Conclusão

---

- Há muito o que ser feito:
  - Técnicas de preparação de dados.
  - Projeto dos algoritmos de aprendizado:
    - Escolha de algoritmos, definição de parâmetros,...
  - Aplicações.



# Conteúdo programático da disciplina

---

- Fundamentos, tipos e tratamento de dados.
- Tipos de Aprendizado: aprendizado supervisionado, não-supervisionado, por reforço, semi-supervisionado.
- Tarefas de aprendizado: classificação, regressão, agrupamento e associação.
- Métodos de Aprendizado: baseados em distância, bayesiano, baseados em procura, baseados em otimização, inspirados pela natureza.
- Medidas de Avaliação: amostragem e medidas de desempenho.
- Aplicações.





# Bibliografia

---

- ALPAYDIN, E. Introduction to Machine Learning. MIT Press, 2014. 640p.
- BISHOP, C. M. Pattern Recognition and Machine Learning. Springer, 2016. 738p.
- CASTRO, L. N. Fundamentals of Natural Computing: Basic Concepts, Algorithms, And Applications. CRC Press, 2006. 696p.
- CHAPELLE, O.; SCHÖLKOPF, B.; ZIEN, A. Semi-supervised Learning. MIT Press, 2006. 508p.
- DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern Classification. John Wiley & Sons, 2012. 680p.
- **FACELI, K.; LORENA, A.C.; GAMA, J.; CARVALHO, A.C.P.L.F. Inteligência Artificial: uma abordagem de aprendizado de máquina. LTC, 2011, 394p.**
- GRUS, J. Data Science from Scratch: First Principles with Python. O'Reilly Media, Inc., 2015. 330p.
- HAYKIN, S. O. Neural Networks and Learning Machines. Pearson Education, 2011. 936p.
- KUNCHEVA, L. I. Combining Pattern Classifiers: Methods and Algorithms. John Wiley & Sons, 2014. 384p.
- MITCHELL, T. Machine Learning. McGrawHill, 1997, 414p.
- MURPHY, K. P. Machine Learning: A Probabilistic Perspective. The MIT Press, 2012. 1067p.
- **SILVA, L. N. C.; FERRARI, D. G. Introdução À Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações. Editora Saraiva, 2016, 376p.**
- SILVA, T. C.; ZHAO, L. Machine Learning in Complex Networks. Springer, 2016. 331p.



# Recursos necessários

---

- Conhecimentos em Python e pacotes associados.
- Anaconda
- Disposição para programar.



# Avaliação da disciplina

---

- Trabalhos extra-classes.
  - Tarefas em aulas.
  - Prova.
  - Projeto final.
- 
- Cálculo da média será definido em breve.



FIM



# Inferência Indutiva (1/2)

---

## ■ **Indução**

- Um processo de raciocínio para uma conclusão sobre todos os membros de uma classe por meio do exame de apenas uns poucos membros da classe
- De maneira geral, raciocínio do particular para o geral
- Por exemplo, se eu noto que:
  - Todos os pacientes com Déficit de Atenção atendidos em 1986 sofriam de Ansiedade
  - Todos os pacientes com Déficit de Atenção atendidos em 1987 sofriam de Ansiedade
  - ...
  - **Posso inferir logicamente que Todos os pacientes que sofrem de Déficit de Atenção também sofrem de Ansiedade**
  - Isto pode ser ou não verdade, mas propicia uma boa **generalização**



# Inferência Indutiva (2/2)

---

- De uma maneira mais “formal”...
  - Para um conjunto de objetos,  $X=\{a,b,c,d,...\}$ , se a propriedade  $P$  é verdade para  $a$ , e se  $P$  é verdade para  $b$ , e se  $P$  é verdade para  $c$ ,... então  $P$  é verdade para todo  $X$
  - O conhecimento novo baseado em vários casos (indução) é geralmente verdadeiro desde que os sistemas estudados sejam bem comportados
    - Se o número de objetos (exemplos) for insuficiente, ou se não forem bem escolhidos, as hipóteses obtidas podem ser de pouco valor
  - A inferência indutiva é um dos principais métodos utilizados para derivar conhecimento novo e prever eventos futuros



# Aprendizado de Máquina - uma definição

---

Um programa aprende a partir da experiência **E**, em relação a uma classe de tarefas **T**, com medida de desempenho **P**, se seu desempenho em **T**, medido por **P**, melhora com **E**

*Mitchell, 1997*

Também chamado de **Aprendizado Indutivo**

- Melhorar a realização de uma tarefa a partir da experiência
  - Melhorar a realização da tarefa  $T$
  - Em relação a uma medida de desempenho  $P$
  - Baseada na experiência  $E$





## Aprendizado de Máquina - Exemplo (1/2)

---

- Detecção de bons clientes para um cartão de crédito
  - **Tarefa T:** classificar potenciais novos clientes como bons ou maus pagadores
  - **Medida de Desempenho P:** porcentagem de clientes classificados corretamente
  - **Experiência de Treinamento E:** uma base de dados histórica em que os clientes já conhecidos são previamente classificados como bons ou maus pagadores



## Aprendizado de Máquina - Exemplo (2/2)

---

- Reconhecimento de caracteres manuscritos
  - **Tarefa T**: reconhecer e classificar caracteres manuscritos
  - **Medida de Desempenho P**: percentagem de caracteres classificados corretamente
  - **Experiência de Treinamento E**: base de dados de caracteres manuscritos com a respectiva classificação



# Tipos de Aprendizado de Máquina (1/3)

---

- Aprendizado Supervisionado

- O algoritmo de aprendizado (indutor) recebe um conjunto de exemplos de treinamento para os quais os rótulos da classe associada são conhecidos
- Cada exemplo (instância ou padrão) é descrito por um vetor de valores (atributos) e pelo rótulo da classe associada
- O objetivo do indutor é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados
- Para rótulos de classe discretos, esse problema é chamado de classificação e para valores contínuos como regressão



## Tipos de Aprendizado de Máquina (2/3)

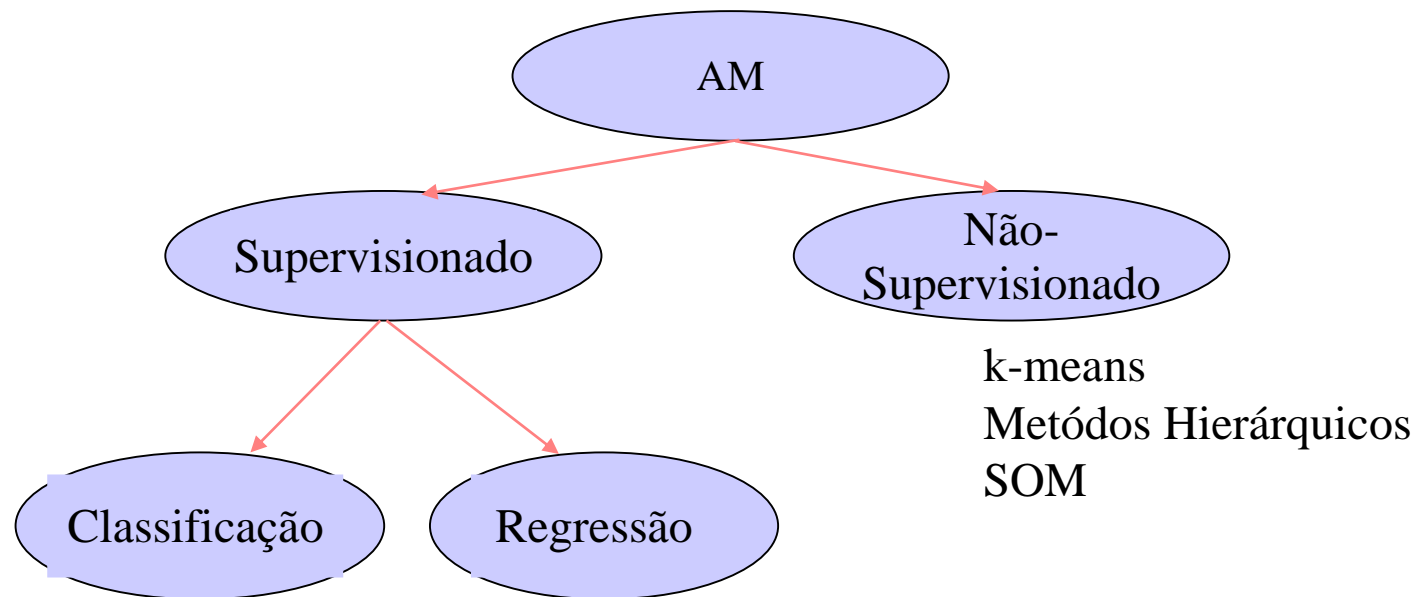
---

- Aprendizado Não-Supervisionado
  - O indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando agrupamentos ou clusters
  - Após a determinação dos agrupamentos, em geral, é necessário uma análise para determinar o que cada agrupamento significa no contexto problema sendo analisado



# Tipos de Aprendizado de Máquina (3/3)

---



k-NN  
Árvores de Decisão  
Naive Bayes  
Perceptron/Adaline  
Multi-Layer Perceptron

k-NN  
Adaline  
Multi-Layer Perceptron

k-means  
Métodos Hierárquicos  
SOM



# Paradigmas de Aprendizado

---

- Simbólico
  - Expressão lógica, árvores de decisão, regras ou rede semântica.
- Estatístico
  - Aprendizado Bayesiano
- Baseado em Exemplos
  - Nearest Neighbours e raciocínio baseado em casos
- Conexionista
  - Redes neurais
- Evolutivo
  - Algoritmos genéticos



## *Conceitos e Definições de Aprendizado*



# Conceitos e Definições

---

- Exemplo (caso, registro ou dado) é uma tupla de valores de atributos
  - Um paciente, dados médicos sobre uma determinada doença
- Atributo: descreve uma característica ou um aspecto de um exemplo.
  - Nominal: cor
  - Contínuo: peso
- Classe: atributo especial (aprendizado supervisionado), denominado rótulo ou classe.
  - Classificação: discreto ou nominal  $C_1, C_2, \dots C_k$
  - Regressão: valores reais





# Conceitos e Definições

---

- Conjunto de exemplos: Um conjunto de exemplos é composto por exemplos contendo valores de atributos bem como a classe associada.

<b>Tempo</b>	<b>Temperatura</b>	<b>Humidade</b>	<b>vento</b>	<b>Joga</b>
Sol	85	85	Não	Não
Sol	80	90	Sim	Não
Nublado	83	86	Não	Sim
Chuva	70	96	Não	Sim
Chuva	68	80	Não	Sim
Chuva	65	70	Sim	Não
Nublado	64	65	Sim	Sim
Sol	72	95	Não	Não
Sol	69	70	Não	Sim
Chuva	75	80	Não	Sim
Sol	75	70	Sim	Sim
Nublado	72	90	Sim	Sim
Nublado	81	75	Não	Sim
Chuva	71	91	Sim	Não



# Conceitos e Definições

---

- Classificador ou Hipótese: Dado um conjunto de exemplos de treinamento, um indutor gera como saída um classificador (também denominado hipótese ou descrição de conceito) de forma que, dado um novo exemplo, ele possa predizer com a maior precisão possível sua classe.
- Ruído: é comum no mundo real, trabalhar com dados imperfeitos. Eles podem ser derivados do próprio processo que gerou os dados, do processo de aquisição de dados, do processo de transformação ou mesmo devido a classes rotuladas incorretamente (por exemplo, exemplos com os mesmos valores de atributos mas com classes diferentes).



# Conceitos e Definições

---

- Missing Values (Valores Perdidos): em geral, indicados por valores fora do escopo.
  - Tipos: desconhecidos, não registrados, irrelevantes.
  - Razões:
    - Mau-funcionamento do equipamento.
    - Mudanças na definição do experiment.
    - Incapacidade de mensuração.
- Valores perdidos podem, de fato, significar alguma coisa:
  - A maioria dos métodos de aprendizado não assumem isto.
  - No entanto, este tipo de informação pode ser codificado como um valor adicional.



# Conceitos e Definições

---

- Modo de aprendizado:
  - Não incremental (batch): sempre que todo o conjunto de treinamento deva estar presente para o aprendizado.
  - Incremental: o indutor apenas tenta atualizar a hipótese antiga sempre que novos exemplos são adicionados ao conjunto de treinamento.
- Taxa de Erro de um classificador  $h$ :

$$\text{err}(h) = \frac{1}{n} \sum_{i=1}^n \| y_i \neq h(x_i) \|$$

- Compara a classe da exemplo com o rótulo atribuído pelo classificador induzido.



# Conceitos e Definições

---

- Precisão do Classificador: complemento da taxa de erro denotado por  $\text{acc}(h)$

$$\text{acc}(h) = 1 - \text{err}(h)$$



# Conceitos e Definições

---

- Erro de Regressão: calcula-se a distância entre o valor real com o atribuído pela hipótese induzida. Pode-se usar: erro médio quadrado ou a distância absoluta média.

$$\text{mse-err}(h) = \frac{1}{n} \sum_{i=1}^n (y_i - h(x_i))^2$$

$$\text{mad-err}(h) = \frac{1}{n} \sum_{i=1}^n |y_i - h(x_i)|$$